ELSEVIER

Symptom Severity at Week 4 of Cognitive-Behavior Therapy Predicts Depression Remission

Jacqueline B. Persons

Oakland Cognitive Behavior Therapy Center and University of California, Berkeley

Cannon Thomas

San Francisco Group for Evidence-Based Psychotherapy and University of California, San Francisco

Early response has been shown to predict psychotherapy outcome. We examined the strength of the relationship between early response and remission in 82 patients who received naturalistic cognitive-behavior therapy in a private practice setting, and 158 patients who received protocol cognitive therapy in a research setting. We predicted that the relationship between early response and remission would be substantial enough to guide clinical decision making in both samples, and that a simple model of severity at Week 4 of treatment would predict remission as effectively as a more complex change score. Logistic regressions showed that a simple model based on the Week 4 Beck Depression Inventory (BDI) score was as predictive of remission as more complex models of early change. A receiver operating characteristics analysis showed that BDI score at Week 4 was substantially predictive of remission in both the naturalistic and research protocol samples; the area under the curve was .80 and .84 in the naturalistic and protocol samples, respectively. To guide clinical decision making, we identified threshold scores on the BDI corresponding to various negative predictive values (probability of nonremission when nonremission is predicted). Our results indicate that depressed patients who remain severely depressed at Week 4 of cognitive therapy are unlikely to reach remission at the end of relatively brief (maximum 20 sessions) treatment. We discuss implications of our findings for clinical decision making and treatment development.

THE PSYCHOTHERAPIST WHOSE PATIENT IS NOT DOING WELL frequently confronts the question: Will my patient ultimately benefit from the treatment I am providing (and thus I should stay the course), or is he or she unlikely to benefit (and thus I should change or end the treatment)? The answer to this question has important implications. Ethics are at stake: The American Psychological Association (2002) ethics code proscribes psychologists from providing unhelpful treatment. In addition, if the therapist persists too long with an unhelpful treatment or prematurely ends a treatment that would eventually have succeeded, the patient's suffering is prolonged and resources are wasted.

The importance of the psychotherapist's question is highlighted by the fact that large numbers of patients fail to benefit from psychotherapy, even

from empirically supported treatment (EST; Cuijpers et al., 2014), and by evidence that clinicians' decision making is often poor (Garb, 2005), especially when the patient is not improving (Kendall, Kipnis, & Otto-Salaj, 1992; Stewart & Chambless, 2008). Information about predictors of outcome, and assistance to help the clinician use this information to guide decision making, are sorely needed.

We know that most symptom change happens early in psychotherapy. The review by Ilardi and Craighead (1994) of eight studies of cognitive-behavioral therapy (CBT) for depression showed "60%–80% of the total decrease in depression severity typically occurring by Week 4" (p. 142). We also know that patients who report a substantial decrease in symptoms early in treatment have a better outcome of CBT for depression (Crits-Christoph et al., 2001; Fennell & Teasdale, 1987; Schindler, Hiller, & Witthoft, 2013; Steidtmann et al., 2013), panic disorder (Aaronson et al., 2008; Lutz et al., 2014), problem drinking (Breslin, Sobell, Sobell, Buchan, & Cunningham, 1997), binge-eating disorder (Grilo, White, Masheb, & Gueorguieva, 2015; Grilo, White, Wilson, Gueorguieva, & Masheb, 2012), and bulimia nervosa (Fairburn, Agras, Walsh, Wilson, & Stice, 2004) than patients who do not show an early response.

Does it make sense for clinicians to use these published findings as the basis for making a change in the treatment plan when the patient fails to show an early response? The reliability and robustness of the relationship between early response and treatment outcome suggest that the answer to this question is "yes."

However, the therapist who is striving to do EST hesitates to deviate from the EST protocol. The EST protocol for cognitive therapy for depression (e.g., Beck, Rush, Shaw, & Emery, 1979), for example, calls for a sequence of interventions delivered in 18–20 sessions over 16 weeks, and the therapist who adheres to the manual will deliver the complete protocol without modification.

What are clinicians to do? When our patient fails to show an early response to treatment, should we attend to the evidence that indicates the patient is unlikely to benefit much from treatment? If so, at that point, we'd discuss the issue with our patient and consider making a change in the treatment plan, or should we attend to the evidence from the randomized controlled trials that tells us that the protocol, as written, provides effective treatment? If so, at that point we would simply follow the protocol as written. In the study reported here, we addressed the question of whether the relationship between early response and outcome of CBT for depression is strong enough to guide clinical decision making. We

did that by conducting receiver operating characteristic (ROC) analyses to examine the strength of the relationship between early response and remission at the end of treatment in patients receiving CBT for depression in naturalistic and research settings.

We also asked whether the effect of early response on outcome is more related to the fact that patients who have had an early response have experienced a large *change* during the early stages of treatment or to the fact that they are *less severely ill* after the early treatment period. Without the answer to this question, the clinician doesn't know which of these two pieces of information to attend to. In addition, if information about *severity* after early treatment was sufficient, it would provide a simpler heuristic (a cutoff score) than the *change* score, which requires the clinician to make a calculation. Based on past research and our clinical experience, we predicted that early *change* was more predictive of outcome than early *severity*.

Finally, to guide clinical decision making, we used the data from the samples we studied to identify threshold scores on the Beck Depression Inventory (BDI; Beck, Ward, Mendelsohn, Mock, & Erbaugh, 1961) corresponding to various negative predictive values (probability of nonremission when nonremission is predicted). We focused on negative predictive value because, following Lambert (2010), we viewed early identification of the nonresponder to be the clinician's highest priority prediction task.

To strengthen the reliability and generalizability of our results, we studied seven samples of patients who received cognitive therapy for depression—one naturalistic sample of patients treated in a private practice setting, and six samples of patients who received protocol-guided treatment provided in a research setting.

Method

THE NATURALISTIC SAMPLE
Selection Criteria

Subjects for this study were selected from archival data obtained from patients who were treated at the San Francisco Bay Area Center for Cognitive Therapy, a group private practice in Oakland, California, or by JBP before she established the practice. All patients who contributed data to the database gave written permission for use of anonymized data for research purposes in the treatment agreement they signed at the beginning of treatment.

We selected cases for study if they (a) were ages 18 or over, (b) received individual therapy, (c) completed a BDI at the intake session, (d) had a BDI score of 20 or greater at intake, (e) provided a BDI score at Week 4

(28 +/– 3 days postintake), and (f) provided at least one BDI score after Week 4. We used minimal selection criteria in order to maximize our sample size and capitalize on the naturalistic nature of the sample. We selected adults receiving individual therapy. We selected those who provided a BDI score at the intake session of 20 or greater in order to select a sample that was comparable to the patients treated in the randomized controlled trials, which used this selection criterion. We selected patients who provided a BDI score within Week 4 of treatment (28 days +/– 3 days), so that we could evaluate whether the patient was an early responder. (We provide a rationale below for selecting Week 4 as the early response point.) We required patients to have at least one BDI score after Week 4 so we could assess the patient's treatment outcome. The sample that met these criteria consisted of 82 patients.[1]

In the naturalistic sample, of the total sample of 1,436 patients in the data set, 966 (67.3%) provided an intake BDI score. Of those, 325 (33.6%) met the selection criterion of BDI ≥ 20. Of those, 155 remained in therapy for 5 or more weeks (47.7%; we selected patients who stayed in treatment 5 weeks or more because we needed one more BDI score after Week 4 to serve as the termination BDI score). Of those 155 patients, 82 (52.9%) had a BDI score at Week 4.

Treatment

Treatment consisted of CBT, usually provided weekly in the beginning and middle phases of treatment, and tapering to every other week or monthly near the end of treatment. Therapists did not adhere to a manualized treatment. Instead, they developed an individualized cognitive-behavioral case formulation and used the formulation and the results of progress monitoring data to make treatment decisions, including to select interventions from the available CBT manuals and other sources (Persons, 2008). Patients completed a BDI in the waiting room before the session, and the therapist typically plotted the score at the beginning of the session, reviewed the plot with the patient, and used the data to guide decision making in the session. Therapists could recommend a wide range of changes to the treatment to address the patient's needs, including adding modalities (e.g., recommending pharmacotherapy or other adjunctive treatment, such as couple therapy), including the patient's significant others in therapy sessions, changing the cognitive-behavioral model that was

used to formulate and intervene (e.g., shifting from the Beck et al., 1979 cognitive model to a behavioral [e.g., Lewinsohn, Gotlib, & Hautzinger, 1998] or behavioral activation [e.g., Martell, Addis, & Jacobson, 2001] model of depressive symptoms), or targeting a comorbid difficulty (e.g., an anxiety disorder or attention-deficit/hyperactivity disorder [ADHD]). Sixty-eight percent of patients received pharmacotherapy at some point during their treatment and 19% received some other type of adjunctive treatment.

Patients were treated by JBP or by one of 17 other therapists at the practice; most therapists were doctoral-level psychologists, one had a master's in social work. The mean duration of treatment completed by patients in the sample was 26.8 weeks (SD = 24.6); the median number of weeks in treatment was 16. Treatment was open-ended in duration and ideally ended when patient and therapist agreed that the patient had reached the patient's goals. Sometimes treatment ended for other reasons, including the therapist leaving the practice, the patient moving away or dropping out of treatment, or (rarely) when the therapist concluded that he or she was unable to be helpful and referred the patient to another provider.

THE PROTOCOL SAMPLES

We selected samples of patients who received protocol treatment in randomized controlled trials of CBT who provided weekly scores on the BDI during the course of treatment. We obtained data from six samples: the cognitive therapy condition of the Treatment of Depression Collaborative Research Program (TDCRP; Elkin et al., 1989), all three conditions of CBT (behavioral activation; behavioral activation plus interventions targeting automatic thoughts; and treatment that targeted behaviors, automatic thoughts, and schemas) in the dismantling study conducted by Jacobson et al. (1996), and the cognitive therapy condition and the functional analytic psychotherapy-enhanced cognitive therapy condition (FECT) of Kohlenberg, Kanter, Bolling, Parker, and Tsai (2002), for a total of 158 patients treated in six treatment conditions in three randomized controlled trials.[2]

Selection Criteria

Patients in the protocol samples were included if they had a BDI score immediately prior to the first treatment session of at least 20, a BDI score at Week 4, and at least one BDI score after Week 4.

---

[1] Pre- and posttreatment BDI scores for some of these subjects were presented in four previously published articles (Persons, Bostrom, & Bertagnolli, 1999; Persons, Burns, & Perloff, 1988; Persons, Roberts, Zalecki, & Brechwald, 2006; Thomas & Persons, 2013).

[2] All of these studies were randomized controlled trials except for the Kohlenberg, Kanter, Bolling, Parker, and Tsai (2002) trial, which did randomly assign patients to treatment conditions.

In the total protocol sample of 249 patients, 245 (98%) provided a BDI score at Session 1. Because a time interval elapsed between the participant meeting the BDI ≥ 20 criterion at screening and the first therapy session, 182 (74%) of the 245 participants in the protocol sample met the pretreatment cutoff of BDI ≥ 20 in Session 1. Of those, 163 (90%) remained in treatment for 5 weeks or more. Of those 163 patients, 158 (97%) completed a BDI at Week 4.

Treatment

Treatment consisted of some type of manualized CBT. All of the protocol therapies were designed to consist of 20 sessions of treatment over 16 weeks, with the first 8 sessions occurring twice-weekly over 4 weeks, followed by 12 sessions of weekly treatment.

SYMPTOM MEASURE AND COLLECTION PROCEDURE

Symptoms of depression were assessed in all samples with the original version of the BDI (Beck et al., 1961). The BDI is a widely-used, 21-item self-report measure of the severity of depressive symptoms. Beck, Steer, and Garbin (1988) reported a test–retest reliability over time periods less than 1 month of .60, and Clark and Watson (1991) reported good convergence with other self-report measures of depressive symptoms.

In the naturalistic sample, the patient's intake BDI score was collected before the first meeting with the therapist and at the beginning of every therapy session. Treatment outcome was assessed with the last BDI score available in the patient's clinical record for that episode of treatment. A treatment episode was deemed ended when the clinical record indicated that the case was closed or transferred to another therapist at the practice, or an interval of 6 months or more occurred between sessions.

In the protocol samples, BDI scores were collected during an intake evaluation, at every treatment session, and at other scheduled assessment points during and following the treatment. We studied only the BDI scores collected in the treatment sessions, to conform most closely to the naturalistic sample. The last BDI score we studied was the one collected in the final therapy session.

EARLY RESPONSE PERIOD

We defined the period for early response on the basis of the change in the patient's BDI score between the initial BDI (completed immediately before the first session of treatment) and Week 4 of treatment (completed 28 days +/– 3 days after initial session). We examined week of treatment rather than session in order to match the metric used by Ilardi and Craighead (1994) and many other studies of this topic. We selected Week 4 because that number matched earlier studies of this topic (Fairburn et al., 2004; Grilo et al., 2015; Ilardi & Craighead, 1994) and because a plot of average BDI scores at each treatment session for a sample of patients from the private practice sample that we examined for this purpose (see Figure 1) showed an inflection at the Week 4 point, with more rapid improvement prior to Week 4.

DEFINITION OF REMISSION

We defined remission as a BDI score at the end of treatment of less than 10, as this was the definition used in the TDCRP (Elkin et al., 1989).

Results

We report descriptive statistics for all samples in Table 1. The table shows, for each sample, the number in the sample, mean initial BDI score, mean Week 4 BDI score, mean posttreatment BDI score, percentage of the sample who reached remission, median number of weeks in treatment, and median number of sessions of treatment. The most notable distinction between the samples is that the patients in the naturalistic sample received fewer sessions of treatment before and after Week 4 than the patients in the protocol samples. Patients in the naturalistic sample generally received weekly sessions early in treatment, often tapering to less frequent sessions as treatment progressed. In contrast, patients in the protocol samples received twice-weekly sessions in the first 4 weeks, followed by weekly sessions for the rest of treatment.

Overview of Analyses

We conducted logistic regressions with ROC curve analyses predicting patients' status as remitted or nonremitted at the end of treatment. The first issue we addressed was, What is the simplest information necessary for optimal prediction of end-of-treatment remission status in each sample? We answered this question by estimating three different logistic regression models for each sample in order to identify the one that optimally predicted remission status. The first model we examined was the full model, which included information about severity of symptoms at the initial session, severity at Week 4, and the rate of change between the two time points.[3] We examined whether all of this information (the full model) was

---

[3] The rate of change was calculated as intake BDI score minus BDI score at Week 4, divided by intake BDI score. The statistical limitations of change scores are well established (e.g., see Castro-Schilo & Grimm, 2018), and we could have used alternative methods, such as residualized change scores. We elected to use simple change scores because this is the method for assessing change that the average clinician can use. Clinicians are not likely to have access to the tools and skills needed to calculate more complex change metrics.
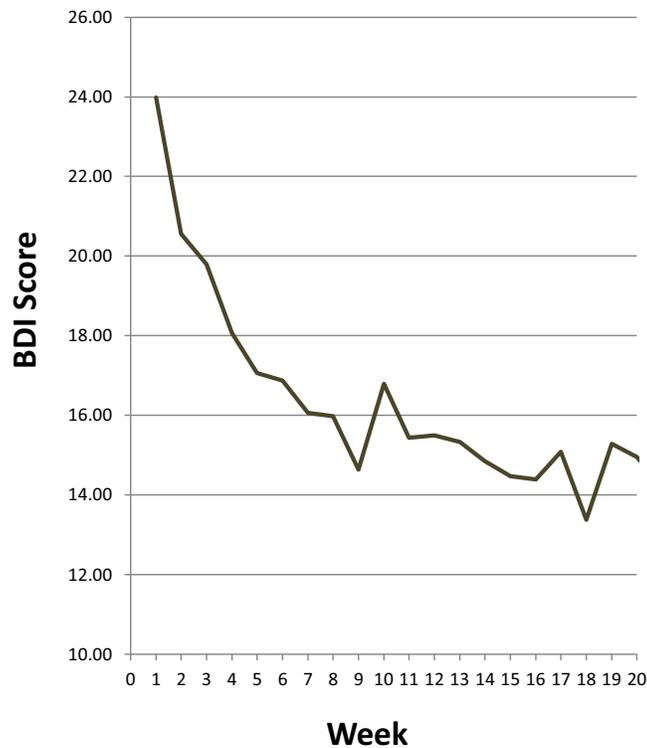
FIGURE 1   BDI score at each week in treatment (*N* = 268).

necessary to optimize prediction or whether a reduced model, including either (a) symptom severity at Week 4 alone or (b) change over the 4 weeks alone, was as predictive as the full model. We used the most parsimonious model that did not significantly reduce prediction for subsequent analyses.

Table 1
Descriptive Statistics for the Naturalistic and Protocol Samples

	N	Initial BDI mean (*SD*)	Week 4 BDI mean (*SD*)	Post-Tx BDI mean (*SD*)	Percentage remission	Median weeks of Tx	Median sessions of Tx [a]
Naturalistic sample	82	27.5 (7.1)	19.4 (10.7)	13.7 (11.7)	50.0	16	11.5 (3/9)
TDCRP CT	35	31.0 (5.8)	23.1 (11.5)	13.5 (13.0)	54.3	16	20 (8/12)
Jacobson BA + AT + CT	38	27.6 (5.7)	19.1 (11.0)	12.0 (10.6)	50.0	16	20 (8/12)
Jacobson BA + AT	30	26.8 (5.8)	18.2 (8.4)	10.4 (9.1)	70.0	16	20 (8/12)
Jacobson BA	33	26.5 (4.9)	17.7 (10.6)	9.9 (8.8)	63.6	16	20 (8/12)
Kohlenberg CT	7	28.7 (4.9)	20.7 (12.1)	14.7 (9.6)	57.1	16	16 (8/12)
Kohlenberg FECT	15	24.7 (3.7)	15.7 (6.5)	9.7 (5.6)	33.3	16	20 (8/12)

Note. BDI = Beck Depression Inventory; TDCRP CT = cognitive therapy condition of Treatment of Depression Collaborative Research Program (Elkin et al., 1989); Jacobson BA + AT + CT = behavioral activation + automatic thoughts + schema change condition of Jacobson et al. (1996); Jacobson BA + AT = behavioral activation + automatic thoughts condition of Jacobson et al. (1996); Jacobson BA = behavioral activation condition of Jacobson et al. (1996); Kohlenberg CT = cognitive therapy condition of Kohlenberg, Kanter, Bolling, Parker, and Tsai (2002); Kohlenberg FECT = functional analytic psychotherapy—enhanced CT condition of Kohlenberg et al. (2002).
[a] The parentheses represent median sessions in first four weeks/median sessions after first four weeks.

Next we asked whether the predictability of final outcome by the logistic regression model was substantial enough to be useful for clinical decision making. We did this by assessing the magnitude of the area under the ROC curve for the model. We used the Hosmer and Lemeshow (2000) Goodness-of-Fit Test to assess the appropriateness of all models tested. For all models, the Hosmer-Lemeshow statistic was nonsignificant, indicating no clear violations of model fit.

Finally, to aid in clinical decision making, we calculated threshold scores at which a patient had a given probability of failing to remit. These scores were simply the scores at which the negative predictive value (probability of nonremission when nonremission is predicted) was 70, 80, or 90%. We selected these three values because they were high enough to provide clinically useful information to help the practitioner identify patients who were not on track to remit.

A full spectrum of efficiency statistics (e.g., negative predictive value, positive predictive value, specificity, sensitivity, Youden's index) is available for evaluating the results of ROC analyses. We made a purposeful choice to calculate and present negative predictive values as guides to the clinician's decision making, rather than the common approach of identifying a cut point that maximized both the sensitivity and specificity of the ROC analyses (e.g., Coffman, Martell, Dimidjian, Gallop, & Hollon, 2007). We focused on the negative predictive value because, following Lambert (2010) and based on our own experience of the clinician's needs as highlighted in the opening paragraph of our article, we viewed early identification of the nonresponder to be the clinician's highest priority prediction task.

PROTOCOL SAMPLE AGGREGATION

Because of the relatively small size of some of our protocol samples for the purposes of logistic regression, we conducted power analyses to evaluate the N necessary to reliably identify effects of the magnitude we expected. We based our analyses on observed values in our naturalistic sample in the optimal model we ended up selecting below: the predictor variable was assigned a mean of 19, SD = 10, and the response probability was .5. Alpha level was set to .05 and power to .90 in order to reliably identify a small-to-moderate-size effect (odds ratio = 1.10 per unit decrease in the predictor variable). An odds ratio of 1.10 is slightly smaller than the observed effect in predicting remission status with Week 4 BDI in the naturalistic sample. Our calculation showed that a sample size of 65 was required. Therefore, in order to maximize our power in detecting effects of interest in this study, we aggregated the protocol samples into one combined protocol sample.

In order to justify sample aggregation, we needed to demonstrate that differences between samples did not affect results. Consequently, we examined whether a fixed effect variable that coded for original sample membership made a statistically significant difference in the fit of the optimal model that we ended up selecting below. We found that both the fixed effect nested likelihood ratio test (ΔLRT = 14.7(5), p = .01) and the interaction with the other predictor (ΔLRT = 12.9(5), p = .02) significantly improved the fit of the model, indicating that differences between the samples did affect early prediction of remission. Follow-up analyses indicated that the sample effect was due to a single sample, the FECT condition of Kohlenberg et al. (2002). When we excluded this sample from the combined sample, neither the fixed effect for sample (ΔLRT = 4.1(4)) nor the interaction (ΔLRT = 2.7(4)) was statistically significant. Therefore, we examined results for the aggregated protocol sample both with and without including the FECT sample. We found that the inclusion of the FECT sample did not change any of our conclusions below. Results of the combined sample including and not including the FECT sample are presented in Table 2, so that readers can assess the lack of impact of including the FECT sample for themselves. Given the lack of a practical justification for excluding the FECT sample, all subsequent discussion focuses on the combined sample that includes all protocol samples: the three conditions of the Jacobson et al. (1996) study, the TDCRP sample, and the Kohlenberg et al. (2002) CT and FECT samples. Total N was 158.

SELECTION OF THE OPTIMAL MODEL

For both samples (naturalistic sample and combined protocol sample), we examined the question of what was the minimal information at the 4-week point of treatment that was necessary for optimal prediction of final remission status. We looked at a full model containing the following predictor variables: intake BDI score, Week 4 BDI score, and the rate of change between the intake BDI score and Week 4 BDI score.[4] We then considered each of the variables as an individual predictor. The full model contains redundancy, because one variable is derived from the other two. But its advantages are that (a) it contains all of the information that might be of use in predicting remission status and (b) all single predictor variable models are nested within

---

[4] Interactions between predictor variables were considered in all model series presented in this paper, and the interactions did not contribute to overall model fit in any case. For clarity of presentation, we did not include the details about the interactions here.

Table 2
Fit Indices for Logistic Regression Models Predicting End-of-Treatment Remission Status for the Naturalistic and Combined Protocol Samples

		Naturalistic sample	Combined protocol sample (excluding FECT)	Combined protocol sample (including FECT)
BDI0 + BDI4 + RATE	Wald	16.7(3), < .0001	33.0(3), <.0001	36.4(3), <.0001
	LRT	27.0(3), < .0001	58.4(3), <.0001	61.8(3), <.0001
	ΔLRT	———	—	———
	AUC	.80(.71–.90)	.85(.79–.91)	.84(.78–.90)
BDI0 only	Wald	4.2(1), .04	9.3(1), .002	5.8(1), .02
	LRT	4.7(1), .03	10.0(1), .002	6.0(1), .01
	ΔLRT	22.7(2), <.0001	48.4(2), <.0001	55.8(2), <.0001
	AUC	.66(.54–.78)	.64(.55–.73)	.59(.50–.69)
BDI4 only	Wald	16.4(1), <.0001	33.3(1), <.0001	35.3(1), <.0001
	LRT	25.3(1), .0001	58.0(1), <.0001	59.8(1), <.0001
	ΔLRT	1.7(2), ns	0.4(2), ns	2.0, ns
	AUC	.80(.71–.90)	.85(.78–.91)	.84(.78–.90)
RATE only	Wald	14.4(1), .0001	31.4(1), <.0001	35.7(1), <.0001
	LRT	20.2(1), .0001	48.7(1), <.0001	56.0(1), <.0001
	ΔLRT	6.8(2), .03	9.7(2), .008.	5.8(1), .02
	AUC	.78(.68–.88)	.82(.75–.89)	.83(.76–.89)

Note 1. FECT = functional analytic psychotherapy—enhanced CT condition of Kohlenberg, Kanter, Bolling, Parker, and Tsai (2002); BDI0 = Beck Depression Inventory score at intake; BDI4 = BDI score at 4 weeks; RATE = percentage change in BDI score from intake to Week 4; Wald = Wald statistic; LRT = likelihood ratio test; ΔLRT = change in likelihood ratio test. All statistics are listed with degrees of freedom and *p* values.
Note 2. The change in the likelihood ratio test (ΔLRT) is the difference in the LRT statistic between a model and a simpler model fully nested within it. It is evaluated on the difference in the number of degrees of freedom between the two models. A statistically significant value for the ΔLRT indicates the nested model accounts for significantly less variance than the full model (BDI0 + BDI4 + RATE). AUC = area under the curve (Wald confidence interval).

it. We were particularly interested in whether rate of change alone or Week 4 BDI score alone would have the predictive power of the full model. The full model was used only for purposes of comparison to the reduced models, because high multicollinearity in the full model would have made individual parameter estimates unreliable (Yoo et al., 2014).

Fit statistics and results of ROC curve analyses for each of the models for the naturalistic and combined protocol samples are presented in Table 2. In order to assess model fit, we include the Wald statistic, the likelihood ratio test (LRT), and the ΔLRT. The ΔLRT was our primary statistic of interest, as it allows a determination of whether a statistically significant decrease in the prediction of the criterion variable occurs when predictor variables are removed from the model. The ΔLRT is distributed as a χ^2 evaluated on the number of degrees of freedom difference between the reduced model and the full model. We also calculated the area under the curve (AUC) for the ROC curve created from each model and Wald 95% confidence intervals for the AUC. The AUC provides an indication of whether there are any practical differences in predictive power between the models.

In both the naturalistic and combined protocol samples, the result for the model based on Week 4 BDI score is strongest. The ΔLRT shows that no

information is lost from the full models in the Week 4 BDI models, whereas the rate of change models are statistically significantly weaker than the full models. Moreover, the AUC estimate for the Week 4 BDI models are equal to the full models. All of these results held for the protocol sample whether or not the FECT sample was included.

These findings support the position that a model based on a single variable (BDI score at Week 4) is as predictive of final remission status as more complex models containing information about rate of change up to Week 4. Week 4 BDI score requires no calculation on the clinician's part, and thresholds for determining whether a patient is likely to fail to recover are easier for a clinician to interpret than change scores.

We also examined the predictive power of the BDI score at baseline, and as Table 2 shows, results of the ΔLRT analysis shows that the baseline score alone has statistically significantly less ability to predict remission than does the full model.

PREDICTIVE POWER OF THE WEEK 4 MODEL
Having selected our model (BDI score at Week 4), we now can consider in more detail how practically effective it is at predicting remission. Table 2 presents information about ROC curves based on Week 4 BDI scores for all samples. In the

naturalistic sample, Week 4 BDI is a moderate predictor of end-of-treatment outcome (.80). In the combined protocol sample, the AUC was somewhat higher, with a value of .84. We interpreted results as indicating that Week 4 BDI score is a moderate predictor of remission in both the naturalistic and protocol samples.

THRESHOLD SCORES FOR NEGATIVE PREDICTIVE VALUES

To aid in clinical decision making, we calculated and report in Table 3 threshold scores for the BDI at Week 4 for various negative predictive values (the likelihood a patient above the threshold will fail to remit) for the naturalistic and combined protocol samples.

Results show that patients who have a BDI score at Week 4 of ≥ 25 in the naturalistic private practice sample and ≥ 30 in the combined protocol sample have a 90% probability of failing to remit after a full course of treatment. The thresholds for 70% likelihood of failing to remit are 17 or higher in the naturalistic sample and 20 or higher in the combined protocol sample. The published BDI cutoff scores for depression severity are 19–29 for moderate depression and 30–63 for severe depression (Beck et al., 1988). Our thresholds suggest that patients who remain moderately depressed at Week 4 in the naturalistic sample have greater than a 70% chance of failing to remit, and all patients in the combined protocol sample who remain moderately depressed except those scoring at the very low end of the moderate range have a 70% or greater chance of failing the treatment. In both the naturalistic and combined protocol samples, a patient who is severely depressed at Week 4 has a 90% or greater chance of failing to remit.

Discussion

Results support our hypothesis that the severity of the depressed patient's symptoms at Week 4 of cognitive therapy for depression is a sufficiently strong predictor of remission to guide clinical decision making. The area under the curve for the

Table 3
BDI Cutoff Scores at Week 4 for Various Percentages of Negative Predictive Values (NPV) of Patients' Risk of Failing to Remit

Sample	NPV = 70%	NPV = 80%	NPV = 90%
Naturalistic (N = 82)	≥ 17	≥ 22	≥ 25
Combined protocol (N = 143)	≥ 20	≥ 25	≥ 30

Note. BDI = Beck Depression Inventory.

ROC analyses of remission based on Week 4 BDI score were .80 and .84 in the two samples we studied. These figures indicate that Week 4 BDI score is a moderate predictor of remission following a relatively brief (maximum 20 sessions) cognitive therapy for depression.

We found that a very simple metric, BDI score at Week 4, is just as predictive of end-of-treatment remission status as more complex models that include information about the rate of change to that point. This finding advances the important work of Lambert, Harmon, Slade, Whipple, and Hawkins (2005), who have shown that patients who are treated by clinicians who receive an alert early in treatment to identify patients who are "not on track" for a good outcome have better outcomes than patients whose therapists do not receive an alert. Unfortunately, Lambert et al.'s methods are difficult to apply for clinicians in routine practice who do not have access to Lambert et al.'s software, which relies on benchmarking data from thousands of patients to derive the algorithm used to determine whether or not the patient is "on track." Our results suggest that a clinician may be able to determine whether a patient receiving cognitive therapy for depression is "on track" by consulting a simple threshold BDI score (see Table 3) at Week 4.

The fact that Week 4 BDI score was a moderate predictor of remission means that we were able to calculate negative predictive values (see Table 3) that identify BDI scores at Week 4 above which the patient's likelihood of remission at the end of treatment is quite low. Our results show that patients who have a BDI score of ≥ 25 in the naturalistic private practice sample and ≥ 30 in the pooled protocol sample at Week 4 of treatment have a 90% probability of failing to remit after a full course of treatment. The published BDI cutoff scores for depression severity are 19–29 for moderate depression and 30–63 for severe depression (Beck et al., 1988). Our thresholds suggest that in both the naturalistic and protocol samples, a patient who is severely depressed at Week 4 has a 90% or greater chance of failing to remit. The fact that the finding is consistent across naturalistic and protocol samples, despite substantial differences in the frequency of sessions before and after Week 4, strengthens the generalizability of the finding.

Our data allow us to compare the predictive power of Week 4 BDI score in the naturalistic and protocol samples. We might hope that because clinicians in the naturalistic sample were more able to tailor treatment to address the patient's unique needs, and had access to their client's BDI score at every session and the ability to make changes in the

treatment based on the score, the BDI score at Week 4 might be less predictive of outcome than in the protocol samples, where treatment was fixed and guided by the protocol. However, although the AUC for the ROC analyses of remission based on Week 4 BDI score was a bit smaller for the naturalistic sample (.80) than for the combined protocol sample (.85), the difference was not statistically significant. (Our assessment of significance was based on a crude t score calculated from the Wald estimated standard errors for the AUC in the two groups, their respective sample sizes, and the difference in AUC across groups, $t(223)$ .88, ns.) Thus, although our findings suggest that cognitive therapists can provide better care to their depressed patients if they collect a BDI score at every session and use it to guide their decision making, simply having this information and the potential to make changes in the treatment based on it did not weaken the relationship between symptom severity and remission in the naturalistic sample, as we might have hoped.

Perhaps the therapists in the naturalistic sample were not able to make good use of the outcome data they collected because they did not receive a clear signal, like the one Lambert's outcome tracking system produces, of whether the data they were viewing indicated their patient was on track or off track. Lambert (2016) hypothesized that for therapists to be able to make good use of outcome monitoring data, they needed not just the data but also a clear indicator of whether the score they were viewing was more typical of an on-track or off-track case. The cutoff scores in Table 3 provide this clear signal.

The usefulness of the cutoff scores in Table 3 are limited by the fact that they are based on the original form of the BDI, which has been supplanted by the BDI-II (although its 2-week time frame limits its usefulness for progress monitoring purposes). We used the original form of the BDI because we had the data based on this measure to make the comparisons we were interested in studying. To translate our findings to the BDI-II, the clinician can take advantage of the fact that the cutoff score (see Table 3) for the protocol sample was 30, which is the cutoff score between moderate and severe depression on the original BDI (Beck et al., 1988). On the BDI-II, a score of 29 or greater indicates severe depression (Beck, Steer, & Brown, 1996).

Our findings do not allow us to offer any guidance to clinicians about what action to take when the BDI score immediately after Week 4 (that is, in Session 5) indicates that the patient is very unlikely to remit. There are data (Department of Veterans Affairs, 2016) showing that adding pharmacotherapy when psychotherapy is failing can lead to improved outcome. We did not study that question in our investigation. Despite our inability to tell clinicians what action to take when early outcome is poor, our demonstration that early severity, even in the face of early response, predicts failure to remit can itself be helpful to clinicians. Notably, Lambert and Shimokawa (2011; see Fortney et al., 2016, for a review) showed that simply providing clinicians with the information that their patient was not on track for a successful outcome led to improved outcomes for those patients, even when the clinician did not receive any guidance about what action to take to improve the patient's outcome.

Later work by Lambert and his group (Simon, Lambert, Harris, Busath, & Vazquez, 2012; Slade, Lambert, Harmon, Smart, & Bailey, 2008) showed that giving clinicians a clinical support tool to help the clinician assess and address factors (e.g., the therapeutic alliance, the patient's social support) that might be contributing to the patient's poor outcome led to further improvements in outcome as compared to the outcome of patients whose therapists did not receive the clinical support tool to guide their problem solving. Our findings point to the need for additional work to develop tools to aid clinical decision making when patients are failing to respond to treatment.

The findings reported here must be interpreted in the context of the limitations of our study. Patients in the naturalistic sample did not receive research-quality diagnoses, and they were a homogeneous sample of highly educated mostly Caucasian individuals who were able to afford private care. Naturalistic treatment was not manualized, and differed in content and duration from one patient to another. Our dependent variable was remission, and results would likely be different if we had studied other dependent variables, such as reliable change or treatment response (cf. Cuijpers et al., 2014). The use of the inflection in the naturalistic sample data depicted in Fig. 1 to select Week 4 may have constrained our results. However, we do not view this issue as a critical limitation for two reasons. First, we replicate our finding that Week 4 predicts remission in another sample (the combined protocol sample). Second, we do not claim that Week 4 is better than any other week at predicting outcome. We did not take up the very interesting question of which week is the best week to predict outcome. We simply claim that Week 4 does predict outcome. Finally, our efforts to predict outcome early in treatment are limited by our failure to collect information about each patient's medical and psychiatric comorbidities, life changes during

treatment, and other unique patient factors, including resistance and coping style, all of which have been shown to influence the change process during treatment of depression (Beutler, Clarkin, & Bongar, 2000). Our study also focuses only on outcome as measured by a single self-report scale of symptoms and does not address other important outcomes, including functioning, quality of life, or attainment of idiographic goals.

Our findings are based on data from patients who provided BDI scores in their therapy sessions. Not all patients provided a BDI score for every session, raising the question of whether patients who provided data at Week 4 and end of treatment might differ from those who did not. This issue is particularly salient for the naturalistic sample, which had a much higher rate of missing data than the protocol sample. The fact that the findings we obtained in the protocol sample, where missing data is less of an issue, replicate those we obtained in the naturalistic sample, lends support to the notion that our findings are not likely to result from a selection bias resulting from missing data. However, there are well-established differences between protocol and naturalistic samples, and these samples differed in intervention and data collection procedures, and thus we cannot rule out the possibility that there is a systematic difference between participants with and without missing data.

Our data do not address the question of what outcome might be achieved if treatment duration was extended. Many severely depressed patients may need treatments that are quite a bit longer than the treatments studied here (median of 11.5 sessions in the private practice sample and 16–20 in the research samples). Treatment duration was more variable in the private practice than in the research sample. The variability of treatment duration, and thus the time point at which the final BDI score was collected in the private practice sample can be seen as a limitation of our study, because it introduces variation that may affect the results for that sample. However, this variability can also be seen as a strength, as it is a type of variation that is inherent to a naturalistic sample, and it increases the external validity and generalizability of our findings.

A strength of our investigation is the study of seven samples from three randomized controlled trials and one naturalistic sample of patients who received treatment in a private practice setting. Patients in the naturalistic sample received treatment that was individually tailored to the patient's needs on the basis of an individualized case formulation. The protocol treatments were standardized, and consisted of a range of types of CBT, from the full package of Beck's cognitive therapy to behavioral analysis and FECT. Data reporting on outcomes of patients who receive naturalistic treatment in a private practice setting are rarely published, and the study of clinical decision making in a real-world sample has high external validity. Results were quite consistent across all these samples, and this fact strengthens the reliability and generalizability of our findings.

Our results have direct implications for clinicians. Our data suggest that patients who are severely depressed at Week 4 are not likely to experience a remission of their depressive symptoms. For clinicians who wish to make use of this information in their clinical work, it is important to clarify that the term "at Week 4" actually means after 4 weeks of treatment, and thus refers to the BDI score the clinician collects before Session 5. Thus, for example, the patient who reports a BDI score of 25 or greater at Session 5 has, based on our data (see Table 3), a 90% chance of failing to remit. When the clinician is treating a depressed patient whose symptoms are in the severe range at Week 4, our data indicate that, rather than continuing to provide an EST protocol as written, the clinician instead is advised to evaluate the treatment and consider making a change in the treatment plan.

The finding that early severity is as predictive or more predictive than early change suggests that even if the patient shows change early in treatment, if he or she remains severely ill in Session 5, the prognosis for remission remains poor. It is important for clinicians to be aware that the fact that a patient shows change in the early phase of treatment does not indicate that he or she will show change in the later phase of treatment. These findings support the notion that different change processes drive progress in different states of therapy (Howard, Moras, Brill, Martinovich, & Lutz, 1996).

It is important to state that our findings do not call for clinicians to make important clinical decisions based on BDI scores alone. The clinician will want to make clinical decisions on the basis of many factors in addition to BDI score, such as the patient's treatment goals, life circumstances, medical and psychiatric comorbidities, functioning, and quality of life, as well as factors related to the therapist and the treatment setting.

Our results also have implications for psychotherapy treatment developers. Our findings point to the need for tools and protocols that prompt clinicians to make treatment decisions based on patient progress and other factors and help them do this skillfully. Our data suggest that an algorithm for predicting final outcome that is based on severity following early treatment may do a better

job of predicting outcome than one based on early change. Several types of work along these lines are already under way. One is the work of Lambert and colleagues, described above, to develop a clinical support tool to help clinicians identify and respond helpfully to early signs of treatment failure. Another is the design of treatments like dialectical behavior therapy (Linehan, 1993) that call for the therapist to use information about the patient's symptoms that the patient reports on the diary card to guide clinical decision making in each session. Others include measurement-based care (Fortney et al., 2016), the modular protocols developed by Weisz et al. (2012), and the sequential multiple assignment randomized trial (SMART; Lei, Nahum-Shani, Lynch, Oslin, & Murphy, 2012), which call for clinicians to make intervention decisions based in part on progress monitoring data, as well as other approaches to personalized medicine (e.g., Fisher & Bosley, 2015; Ng & Weisz, 2016).

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

References

Aaronson, C. J., Shear, M. K., Goetz, R. R., Allen, L. B., Barlow, D. H., White, K. S., . . . Gorman, J. M. (2008). Predictors and time course of response among panic disorder patients treated with cognitive-behavioral therapy. *Journal of Clinical Psychiatry*, 69, 418–424. https://doi.org/10.4088/JCP.v69n0312

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.

Beck, A. T., Rush, J. A., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy for depression*. New York, NY: Guilford Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for Beck Depression Inventory–II*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8, 77–100. https://doi.org/10.1016/0272-7358(88)90050-5

Beck, A. T., Ward, C. H., Mendelsohn, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Beutler, L. E., Clarkin, J. F., & Bongar, B. (2000). *Systematic guidelines for treating the depressed patient*. New York, NY: Oxford University Press.

Breslin, F. C., Sobell, M. B., Sobell, L. C., Buchan, G., & Cunningham, J. A. (1997). Toward a stepped care approach to treating problem drinkers: The predictive utility of within-treatment variables and therapist prognostic ratings. *Addiction*, 92, 1479–1489. https://doi.org/10.1111/j.1360-0443.1997.tb02869.x

Castro-Schilo, L., & Grimm, K. J. (2018). Using residualized changes versus difference scores for longitudinal research. *Journal of Social and Personal Relationships*, 35(1), 32–58. https://doi.org/10.1177/0265407517718387

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Evidence and taxonomic implications. *Journal of Abnormal Psychology*, 100, 316–336. https://doi.org/10.1037/0021-843X.100.3.316

Coffman, S. J., Martell, C. R., Dimidjian, S., Gallop, R., & Hollon, S. D. (2007). Extreme nonresponse in cognitive therapy: Can behavioral activation succeed where cognitive therapy fails? *Journal of Consulting and Clinical Psychology*, 75(4), 531–541. https://doi.org/10.1037/0022-006X.75.4.531

Crits-Christoph, P., Connolly, M. B., Gallop, R., Barber, J. P., Tu, X., Gladis, M., & Siqueland, L. (2001). Early improvement during manual-guided cognitive and dynamic psychotherapies predicts 16-week remission status. *Journal of Psychotherapy Practice and Research*, 10, 145–154.

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders*, 159, 118–126. https://doi.org/10.1016/j.jad.2014.02.026

Department of Veterans Affairs. (2016). *VA/DoD clinical practice guideline for the management of major depressive disorder*. Retrieved from https://www.healthquality.va.gov/guidelines/MH/mdd/.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., . . . Parloff, M. B. (1989). NIMH Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry*, 46, 971–982. https://doi.org/10.1001/archpsyc.1989.01810110013002

Fairburn, C. G., Agras, W. S., Walsh, B. T., Wilson, G. T., & Stice, E. (2004). Early change in treatment predicts outcome in bulimia nervosa. *American Journal of Psychiatry*, 161(12), 2322–2324. https://doi.org/10.1176/appi.ajp.161.12.2322

Fennell, M. J. V., & Teasdale, J. D. (1987). Cognitive therapy for depression: Individual differences and the process of change. *Cognitive Therapy and Research*, 11, 253–271. https://doi.org/10.1007/BF01183269

Fisher, A. J., & Bosley, H. G. (2015). Personalized assessment and treatment of depression. *Current Opinion in Psychology*, 4, 67–74. https://doi.org/10.1016/j.copsyc.2015.03.031

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2016). A tipping point for measurement-based care. *Psychiatric Services*, 68(2), 179–188. https://doi.org/10.1176/appi.ps.201500439

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, 1, 67–89. https://doi.org/10.1146/annurev.clinpsy.1.102803.143810

Grilo, C. M., White, M. A., Masheb, R. M., & Gueorguieva, R. (2015). Predicting meaningful outcomes to medication and self-help treatments for binge-eating disorder in primary care: The significance of early rapid response. *Journal of Consulting and Clinical Psychology*, 83(2), 387–394. https://doi.org/10.1037/a0038635

Grilo, C. M., White, M. A., Wilson, G. T., Gueorguieva, R., & Masheb, R. M. (2012). Rapid response predicts 12-month post-treatment outcomes in bing-eating disorder: Theoretical and clinical implications. *Psychological Medicine*, 42(4), 807–817. https://doi.org/10.1017/S0033291711001875

Hosmer, D. W., Jr., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, 51(10), 1059–1064. https://doi.org/10.1037/0003-066X.51.10.1059

Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, 1, 138–156. https://doi.org/10.1111/j.1468-2850.1994.tb00016.x

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*, 295–304. https://doi.org/10.1037/0022-006X.64.2.295

Kendall, P. C., Kipnis, D., & Otto-Salaj, L. (1992). When clients don't progress: Influences on and explanations for lack of therapeutic progress. *Cognitive Therapy and Research*, *16*, 269–281. https://doi.org/10.1007/BF01183281

Kohlenberg, R. J., Kanter, J. W., Bolling, M. Y., Parker, C. R., & Tsai, M. (2002). Enhancing cognitive therapy for depression with functional analytic psychotherapy: Treatment guidelines and empirical findings. *Cognitive and Behavioral Practice*, *9*, 213–229. https://doi.org/10.1016/S1077-7229(02)80051-7

Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice.* Washington, DC: American Psychological Association.

Lambert, M. J. (2016). *Measuring client progress in therapy/ Interviewer: J. B. Persons.* Translating Science to Practice, Society for a Science of Clinical Psychology, http://sscpweb.org/SciPrac.

Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology*, *61*(2), 165–174. https://doi.org/10.1002/jclp.20113

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, *48*(1), 72–79. https://doi.org/10.1037/a0022238

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S. A. (2012). A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology*, *8*, 21–48. https://doi.org/10.1146/annurev-clinpsy-032511-143152

Lewinsohn, P. M., Gotlib, I. H., & Hautzinger, M. (1998). Behavioral treatment of unipolar depression. In V. E. Caballo (Ed.), *International handbook of cognitive and behavioural treatments for psychological disorders.* Oxford, UK: Pergamon/Elsevier Science Ltd.

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder.* New York, NY: Guilford Press.

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., . . . Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology*, *82*(2), 287–297. https://doi.org/10.1037/a0035535

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided action.* New York, NY: Norton.

Ng, M. Y., & Weisz, J. R. (2016). Annual research review: Building a science of personalized intervention for youth mental health. *Journal of Child Psychology and Psychiatry*, *57*(3), 216–236. https://doi.org/10.1111/jcpp.12470

Persons, J. B. (2008). *The case formulation approach to cognitive-behavior therapy.* New York, NY: Guilford Press.

Persons, J. B., Bostrom, A., & Bertagnolli, A. (1999). Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. *Cognitive Therapy and Research*, *23*, 535–548. https://doi.org/10.1023/A:1018724505659

Persons, J. B., Burns, D. D., & Perloff, J. M. (1988). Predictors of dropout and outcome in cognitive therapy for depression in a private practice setting. *Cognitive Therapy and Research*, *12*, 557–575. https://doi.org/10.1007/BF01205010

Persons, J. B., Roberts, N. A., Zalecki, C. A., & Brechwald, W. A. G. (2006). Naturalistic outcome of case formulation-driven cognitive-behavior therapy for anxious depressed outpatients. *Behaviour Research and Therapy*, *44*, 1041–1051. https://doi.org/10.1016/j.brat.2005.08.005

Schindler, A., Hiller, W., & Witthöft, M. (2013). What predicts outcome, response, and drop-out in CBT of depressive adults? A naturalistic study. *Behavioural and Cognitive Psychotherapy*, *41*, 365–370. https://doi.org/10.1017/S1352465812001063

Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure. *Psychotherapy Research*, *22*(6), 638–647. https://doi.org/10.1080/10503307.2012.698918

Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology and Psychotherapy*, *15*, 287–303. https://doi.org/10.1002/cpp.594

Steidtmann, D., Manber, R., Blasey, C., Markowitz, J. C., Klein, D. N., Rothbaum, B. O., . . . Arnow, B. A. (2013). Detecting critical decision points in psychotherapy and psychotherapy + medication for chronic depression. *Journal of Consulting and Clinical Psychology*, *81*(5), 783–792. https://doi.org/10.1037/a0033250

Stewart, R. E., & Chambless, D. L. (2008). Treatment failures in private practice: How do psychologists proceed? *Professional Psychology: Research and Practice*, *39*, 176–181. https://doi.org/10.1037/0735-7028.39.2.176

Thomas, C., & Persons, J. B. (2013). Sudden gains can occur in psychotherapy even when the pattern of change is gradual. *Clinical Psychology: Science and Practice*, *20*, 127–142. https://doi.org/10.1111/cpsp.12029

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., . . . Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth. *Archives of General Psychiatry*, *69*(3), 274–282. https://doi.org/10.1001/archgenpsychiatry.2011.147

Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q. P., & Lillard, J. W., Jr. (2014). A study of effects of multicolinearity in the multivariable analysis. *International Journal of Applied Science and Technology*, *4*(5), 9–19.